

# Μαθηματική Στατιστική

## Εργασία 10

Παρασκευή Πότση  
1112 2016 00180

Γεωργία Μανώλαινα  
1112 2016 00128

Φεβρουάριος 2021

# Contents

1	Υπολογισμοί για την $H_0$	2
2	Υπολογισμοί για την $H_1$	4
3	Λόγος Πιθανοφάνειας & Ακρίβεια	5
4	Η ελπίδα πεθαίνει πάντα τελευταία	6

## Εισαγωγή

Θα ακολουθήσει ένας έλεγχος λόγου πιθανοφάνειας με σκοπό την διαλεύκανση της υπόθεσης ενοχής ή αθώωσης του Μπάμπε από το Μαθηματικό.

Οι δύο υποθέσεις που θα μελετηθούν είναι:

- $H_0$  : το dna που ταυτοποιήθηκε θα μπορούσε να είναι οποιουδήποτε άλλου
- $H_1$  : το dna που ταυτοποιήθηκε είναι του Μπάμπε

Μια γενική ιδέα για τον έλεγχο είναι η εξής:

*Αν το dna προς συζήτηση εμφανίζεται συχνά στον πληθυσμό, τότε το ταυτοποιημένο dna της σκηνής του εγκλήματος θα μπορούσε να προέλθει από οποιονδήποτε άλλο. Αν δεν εμφανίζεται συχνά, ενώ ταιριάζει με το dna του ύποπτου,<sup>1</sup> είναι λιγότερο πιθανό ο ύποπτος να μην είναι ο ένοχος, και να είναι απλά μία σύμπτωση.*

---

<sup>1</sup>εδώ το ταίριασμα είναι η πιθανότητα να προκύπτει από τους γονείς

## Chapter 1

# Υπολογισμοί για την $H_0$

Από τη βάση δεδομένων *data* που μας παρέχεται θα εκτιμηθεί η συχνότητα εμφάνισης του κάθε γονότυπου σε καθένα από τα 5 γονίδια (*loci*) ενδιαφέροντος  $G_1, G_2, G_3, G_4, G_5$ .

Έχουν υπολογιστεί ξεχωριστά οι συχνότητες εμφάνισης  $p_{ij}$  των ομόζυγων γονότυπων τύπου  $(j, j)$  για κάθε γονίδιο  $G_i$ , όπου  $j$  το αντίστοιχο αλληλόμορφο γονίδιο. Για τα  $G_1, G_2 : j \in \{1, 2\}$  γιατί υπάρχουν μόνο δύο δυνατά αλληλόμορφα, ενώ για τα  $G_3, G_4, G_5 : j \in \{1, 2, 3\}$  γιατί αντίστοιχα υπάρχουν τρία αλληλόμορφα.

Επιπλέον έχουν υπολογιστεί και οι συχνότητες εμφάνισης των ετερόζυγων γονότυπων  $p_{ijk}$  όπου  $i \in \{1, 2, 3, 4, 5\}$  δηλώνει το γονίδιο που βρίσκεται, ενώ οι συνδυασμοί των αλληλόμορφων είναι οι εξής:

$$(j, k) \in \begin{cases} \{1, 2\} \times \{1, 2\}, & i = 1, 2 \\ \{1, 2, 3\} \times \{1, 2, 3\}, & i = 3, 4, 5 \end{cases}$$

Η συχνότητα εμφάνισης του dna που ταυτοποιήθηκε στο εργαστήριο, θα προκύψει από το γινόμενο των συχνοτήτων εμφάνισης του κάθε γονιδίου, αφού τα 5 *loci* είναι ανεξάρτητα.

Από τα δεδομένα μας δίνεται ότι ο γονότυπος του dna είναι:

$G_1$	$G_2$	$G_3$	$G_4$	$G_5$
11	22	22	11	33

και με χρήση του περιβάλλοντος της R προέκυψαν τα εξής αποτελέσματα:

$p_{11}$	$p_{22}$	$p_{32}$	$p_{41}$	$p_{53}$
0.6393	0.4965	0.0395	0.2477	0.2472

Table 1.1: Συχνότητες εμφάνισης κάθε γονιδίου στα *data*

Συμπεραίνουμε ότι η ζητούμενη συχνότητα είναι:  $L_0 = 0.000767707$

## Chapter 2

# Υπολογισμοί για την $H_1$

Κάθε γονίδιο διαμορφώνεται από το συνδυασμό δύο αλληλόμορφων που προέρχονται από τους γονείς, το ένα από τη μητέρα και το άλλο από τον πατέρα. Από τα δύο αλληλόμορφα της μητέρας, ο απόγονος κληρονομεί το ένα, με πιθανότητα  $\frac{1}{2}$ . Ομοίως και από τον πατέρα.

Εφόσον λοιπόν έχουμε το dna των γονιών του Μπάμπη, υπολογίζουμε την πιθανότητα ο Μπάμπης να έχει γονότυπο στα 5 *loci* ίδιο με αυτό που ταυτοποιήθηκε στο χώρο του εγκλήματος.

Για παράδειγμα, για το  $G_4$  ο πατέρας είναι  $[1|1]$  και η μητέρα  $[1|2]$ . Εφόσον "θέλουμε" ο Μπάμπης να είναι  $[1|1]$ , ο μόνος τρόπος είναι να κληρονομήσει το πρώτο 1 από τον πατέρα (με πιθανότητα 1) και το δεύτερο 1 από την μητέρα (με πιθανότητα  $\frac{1}{2}$ ). Άρα η πιθανότητα το  $G_4 = [1|1]$  είναι  $1 \cdot \frac{1}{2} = \frac{1}{2}$

Με πολλαπλασιασμό των πιθανοτήτων για κάθε γονίδιο, αφού είναι ανεξάρτητα, συμπεραίνουμε ότι η ζητούμενη πιθανοφάνεια είναι:  $L_1 = 0.0078125$

## Chapter 3

# Λόγος Πιθανοφάνειας & Ακρίβεια

Έχοντας υπολογίσει τις πιθανοφάνειες που αντιστοιχούν στις υποθέσεις μας μπορούμε να υπολογίσουμε και τον λόγο τους.

Ως φίλοι του Μπάμπε αυτό που θα θέλαμε να δούμε είναι ο λόγος  $\frac{L_0}{L_1} > C$ .

Αυτό θα σήμαινε ότι το ενδεχόμενο το ταυτοποιημένο dna να είναι οποιουδήποτε άλλου είναι πιο πιθανό από το να είναι του Μπάμπε, γεγονός που ενισχύει την υπεράσπιση του Μπάμπε.

Ο λόγος που υπολογίζουμε είναι:  $\frac{L_1}{L_0} = 10.17641$

Σύμφωνα και με τον επίσημο Ευρωπαϊκό Οδηγό Εγκληματολογικών Αποτελεσμάτων, οδηγούμαστε στο συμπέρασμα ότι τα στοιχεία που έχουμε στα χέρια μας <sup>1</sup> είναι πιο πιθανά στην περίπτωση που η υπόθεση H που αντιστοιχεί στον αριθμητή είναι αληθής, παρά αυτή που αντιστοιχεί στον παρανομαστή, δηλαδή στην περίπτωση που το ταυτοποιημένο dna ανήκει στον Μπάμπε.

Είναι μάλιστα ένας λόγος της τάξεως του 10, δηλαδή παρέχει μέτρια υποστήριξη για την  $H_1$ .

Μήπως λόγω της μέτριας στήριξης της  $H_1$  έχουμε ελπίδες να υπερασπιστούμε εν τέλει τον φίλο μας Μπάμπε;

---

<sup>1</sup> ταυτοποιημένο dna

## Chapter 4

# Η ελπίδα πεθαίνει πάντα τελευταία

Για να στηρίξουμε περισσότερο τα ευρήματά μας, αποφασίσαμε να δημιουργήσουμε μερικά ακόμα (10000000) τυχαία δείγματα dna (*rdata*) για τα 5 γονίδια που μας ενδιαφέρουν με τις συχνότητες που είχαμε βρει στο 1.1.

Επαναλάβουμε τη διαδικασία παραγωγής τυχαίου δείγματος 5 φορές, ώστε να υπολογίσουμε 5 νέες πιθανοφάνειες (*new\_freq*), δηλαδή συχνότητες εμφάνισης του ταυτοποιημένου dna.

Τέλος, από κάθε πιθανοφάνεια υπολογίσαμε και έναν νέο λόγο πιθανοφάνειας δημιουργώντας το διάνυσμα *new\_LR*.

Τα αποτελέσματα που παίρνουμε δίνονται στον παρακάτω πίνακα:

$LR_1$	$LR_2$	$LR_3$	$LR_4$	$LR_5$
129.9917	119.8236	127.0325	134.0051	127.2394

Table 4.1: Λόγοι Πιθανοφάνειας από τα νέα δεδομένα

Παρατηρούμε ότι με τα νέα περισσότερα δεδομένα οι λόγοι είναι της τάξης του 100, και μάλιστα σε τιμές 120 – 130.

Σύμφωνα με το Οδηγό Εγκληματολογικών Αποτελεσμάτων, όχι μόνο δεν υπερισχύει η  $H_0$ , αλλά η  $H_1$  στηρίζεται ακόμη πιο ισχυρά.

**Συμπέρασμα:** Έχουμε ενδείξεις ότι το ταυτοποιημένο dna είναι πιο πιθανό να ανήκει στον Μπάμπη <sup>1</sup> από ότι είναι να προέρχεται από οποιονδήποτε άλλο.

---

<sup>1</sup>Μαλλον εν τέλει έκρυβε έναν Μπάμπη Σουγιά